Statistics and Probability Letters 79 (2009) 1637–1646

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Duals of random vectors and processes with applications to prediction problems with missing values

Yukio Kasahara [a], Mohsen Pourahmadi [b,*], Akihiko Inoue [c]

[a] Department of Mathematics, Hokkaido University, Sapporo 060-0810, Japan

[b] Department of Statistics, Texas A & M University, College Station, TX 77843-3143, USA

[c] Department of Mathematics, Hiroshima University, Higashi-Hiroshima 739-8526, Japan

ARTICLE INFO

Article history:
Received 8 March 2009
Received in revised form 8 April 2009
Accepted 14 April 2009
Available online 22 April 2009

MSC:
62M20
60G10
60G25

ABSTRACT

Important results in prediction theory dealing with missing values have been obtained traditionally using difficult techniques based on duality in Hilbert spaces of analytic functions [Nakazi, T., 1984. Two problems in prediction theory. Studia Math. 78, 7–14; Miamee, A.G., Pourahmadi, M., 1988. Best approximations in $L_p(\mathrm{d}\mu)$ and prediction problems of Szegö, Kolmogorov, Yaglom, and Nakazi. J. London Math. Soc. 38, 133–145]. We obtain and unify these results using a simple finite-dimensional duality lemma which is essentially an abstraction of a regression property of a multivariate normal random vector (Rao, 1973) or its inverse covariance matrix. The approach reveals the roles of duality and biorthogonality of random vectors in dealing with infinite-dimensional and difficult prediction problems. A novelty of this approach is its reliance on the explicit representation of the prediction error in terms of the data rather than the predictor itself as in the traditional techniques. In particular, we find a new and explicit formula for the dual of the semi-finite process $\{X_t; t \le n\}$ for a fixed n, which does not seem to be possible using the existing techniques.

1. Introduction

Irregular observations, missing values and outliers are common in time series data (Box and Tiao, 1975; Brubacher and Wilson, 1976). A framework for dealing with such anomalies is that of $X = \{X_t\}_{t\in\mathbb{Z}}$ being a mean-zero, weakly stationary stochastic process with the autocovariance function $\gamma = \{\gamma_k\}_{k\in\mathbb{Z}}$ and the spectral density function f, where the problem can be formulated as that of predicting or approximating an unknown value X_0 based on the observed values $\{X_t; t \in S\}$ for a given index set $S \subset \mathbb{Z} \setminus \{0\}$ and the knowledge of the autocovariance of the process. Such a problem is quite important to applications in business, economics, engineering, physical and natural sciences, and belongs to the area of prediction theory of stationary stochastic processes developed by Wiener (1949) and Kolmogorov (1941). By restricting attention to linear predictors and using the least-squares criterion to assess the goodness of predictors, a successful solution seeks to address the following two goals:

(P_1) Express the linear least-squares predictor of X_0, denoted by $\hat{X}_0(S)$, and the prediction error $X_0 - \hat{X}_0(S)$ in terms of the observable $\{X_t; t \in S\}$.

(P_2) Express the prediction error variance $\sigma^2(S) = \sigma^2(f, S) := E|X_0 - \hat{X}_0(S)|^2$ in terms of f.

* Corresponding author.
E-mail address: pourahm@stat.tamu.edu (M. Pourahmadi).

0167-7152/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2009.04.005

The focus in prediction theory has been more on the goal (P_2), and the celebrated Szegö–Kolmogorov–Wiener theorem gives the variance of the one-step ahead prediction error based on the *infinite* past or for the "half-line" index set $S_0 := \{\ldots, -2, -1\}$ by

$$\sigma^2(f, S_0) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda)\mathrm{d}\lambda\right) > 0 \tag{1}$$

if $\log f$ is integrable, and otherwise $\sigma^2(S_0) = 0$. However, when the first n consecutive integers are removed from S_0 or for the index set $S_{-n} := \{\ldots, -n-2, -n-1\}$, $n \geq 0$, the formula for the $(n+1)$-step prediction error variance (Wold, 1954; Kolmogorov, 1941) is

$$\sigma^2(f, S_{-n}) = |b_0|^2 + |b_1|^2 + \cdots + |b_n|^2, \quad n = 0, 1, \ldots, \tag{2}$$

where $\{b_j\}$, the moving average (MA) coefficients of the process, is related to the Fourier coefficients of $\log f$ and $|b_0|^2 = \sigma^2(S_0)$ (see Nakazi and Takahashi (1980), and Pourahmadi (1984); see also Section 3).

A result similar to (1) for the interpolation of a single missing value corresponding to the index set $S_\infty := \mathbb{Z} \setminus \{0\}$ was obtained by Kolmogorov (1941). Specifically, the interpolation error variance is given by

$$\sigma^2(f, S_\infty) = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)^{-1}\mathrm{d}\lambda\right)^{-1} > 0 \tag{3}$$

if $f^{-1} \in L^1 := L^1([-\pi, \pi], \mathrm{d}\lambda/(2\pi))$, and otherwise $\sigma^2(S_\infty) = 0$. The corresponding prediction problem for the smaller index set $S_n := \{\ldots, n-1, n\} \setminus \{0\}$, $n \geq 0$, was stated as open in Rozanov (1967, p. 107) and is perhaps one of the most challenging problems in prediction theory next to (1). The index set S_n is, indeed, of special interest as it forms a bridge connecting S_0 and S_∞; it reduces to S_0 when $n = 0$ and tends to S_∞ as $n \to \infty$. In a remarkable paper, Nakazi (1984) using delicate, but complicated analytical techniques (and assuming that $f^{-1} \in L^1$) showed that

$$\sigma^2(f, S_n) = \left(|a_0|^2 + |a_1|^2 + \cdots + |a_n|^2\right)^{-1}, \quad n = 0, 1, \ldots, \tag{4}$$

where $\{a_j\}$ is related to the autoregressive (AR) parameters of the process (see Section 3). Comparing (2) and (4), it is natural to ask why there is such an "inverse-dual" relationship between them.

Concerning the question above, it is worth noting that Nakazi's technique, if viewed properly, reduces the computation of $\sigma^2(f, S_n)$ to that of the $(n+1)$-step prediction error variance of another stationary process $\{Y_t\}$ with the spectral density function f^{-1}, which we call the *dual* of $\{X_t\}$ (see Definition 1 and Section 4.3). His result and technique have spawned considerable research in this area in the last two decades; see Miamee and Pourahmadi (1988), Miamee (1993), Cheng et al. (1998), Frank and Klotz (2002), Klotz and Riedel (2002) and Bondon (2002). A unifying feature of most of the known results thus far seems to be a fundamental duality principle of the form

$$\sigma^2(f, S) \cdot \sigma^2(f^{-1}, S^c) = 1, \tag{5}$$

where S^c is the complement of S in $\mathbb{Z} \setminus \{0\}$ and $f^{-1} \in L^1$; see Cheng et al. (1998) and Urbanik (2000).

The first occurrence of (5) seems to be in the 1949 Russian version of Yaglom (1963) for the case of deleting finitely many points from S_∞. Proof of (5), in general, like those of the main results in Nakazi (1984), Miamee and Pourahmadi (1988), Cheng et al. (1998), and Urbanik (2000), is long, unintuitive and relies on duality techniques from functional and harmonic analysis and requires $f^{-1} \in L^1$ which is not natural for the index set S_n. Surprisingly, an implicit version of (5) had been developed in Grenander and Rosenblatt (1954, Theorem 1) as the limit of a quadratic form involving Szegö's orthogonal polynomials on the unit circle; see also Simon (2005, p. 165). However, it had remained dormant and not used in the context of prediction theory, except in Pourahmadi (1993).

In this paper, we establish a finite-dimensional duality principle (Lemma 1) which involves the notion of dual of a random vector, and show that some prediction problems, including the above and some new ones, which are related to removing a finite number of indices from S_n and S_∞, can be solved in a unified manner. In Section 2, we present the main lemma, some auxiliary facts about dual of a random vector and their consequences for computing the prediction error variances and predictors. In Section 3, using the lemma we first solve three finite prediction problems for X_0 based on the knowledge of $\{X_t; t \in K\}$ with $K = \{-m, \ldots, n\} \setminus (M \cup \{0\})$, $m, n \geq 0$, where M, the index set of the missing values, is relatively small. Then, we obtain the solutions of Kolmogorov, Nakazi, and Yaglom's prediction problems in a unified manner by studying the limit of the solutions by letting $m \to \infty$, followed by $n \to \infty$. As a consequence, we find a new and explicit formula for the dual of the random process $\{X_t; t \leq n\}$ for a fixed n, which does not seem to be possible using the technique of Urbanik (2000), Klotz and Riedel (2002) and Frank and Klotz (2002). This is particularly useful in developing series representations for predictors and interpolators, and sheds light on the approaches of Bondon (2002) and Salehi (1979). In Section 5, we close the paper with some discussions.